

EPIDEMIC THRESHOLDS ON SCALE-FREE GRAPHS: THE INTERPLAY BETWEEN EXPONENT AND PREFERENTIAL CHOICE

PH.BLANCHARD, C.-H.CHANG, AND T. KRÜGER

ABSTRACT. We study epidemic threshold properties in a scale-free random graph model. We show via a branching process approximation that the divergence of the second moment of the degree distribution is equivalent to the absence of an epidemic threshold. We study further the relation between diameter and epidemic threshold. Absence of an epidemic threshold happens precisely when a positive fraction of the nodes form a cluster of bounded diameter.

1. INTRODUCTION

In the last decade there was an enormous increase of interest in phenomena which show for characteristic observables a power-law behavior. Power-law distributions are often called scale-free distributions due to the absence of characteristic sizes. Most prominent under the huge variety of examples became systems which show so-called selforganized criticality (SOC), [3] [4]. The basic idea to explain the appearance of a power-law-distribution is via the linkage to phase-transitions where the appearance of scale-free structures is known since a long time. Despite many efforts only little is known about the origin of power-law-distributions on a mathematical base. Partially due to the increasing knowledge of the structure of the internet there is a growing number of articles studying the structure of networks with a power law distribution for the degree (see [1] and the references therein).

In this article we want to discuss a specific model of scale-free-graphs based on a partner choice strategy and its epidemic threshold properties. Related questions found recently much interest due to the possible absence of epidemic thresholds for a certain range of power law exponents [8] [9]. We were mainly motivated to our studies by the paper of Pastor-Satorras and Vespignani [9], where on largely heuristic grounds the absence of an epidemic threshold for power-law-graphs with exponents less than 3 was claimed. Since the argumentation in this paper was quite doubtful from a mathematical point of view but somehow intuitively convincing we wanted to study the same question on more solid grounds. To get enough theoretical control we developed a new model of power-law-graphs which seems to be of independent interest. Due to the strong independence properties - as far as this is possible for a scale-free degree-distribution - we were able to carry out a rigorous analysis of certain aspects of the model. For a certain range of parameters we get a power-law distribution for the total degree. For other parameter values there is an interesting domain where one has a fat tail degree distribution with gaps. Here the integrated tail distribution is still of power-law form. In both

Date: 22th May 2003.

cases the divergence of the second moment of the degree distribution is indeed equivalent to the absence of an epidemic threshold. We furthermore examine the relation between divergence of the second moment of the degree distribution and diameter-like properties of the relevant graph-spaces. Namely one has in the case of absence of an asymptotic epidemic threshold the striking property that there is a huge (meaning a positive fraction of the whole population) cluster of finite-size independent-diameter. This clusters carries the epidemics and forms a kind of very small world inside the small world of the largest connected component which can still be of logarithmic diameter.

We further study some models with completely different degree distribution where the absence of an epidemic threshold is not necessary linked to the divergence of the second moment of the degree-distribution.

2. A PREFERENTIAL-CHOICE MODEL OF SCALE-FREE GRAPHS

In the following we want to describe the basic structure of the model we consider in this paper. Our first aim is to construct a family of static random graph models in which vertex degrees are distributed power-law like, while edges still have high degree of independence. As usual in random graph theory we will entirely deal with asymptotic properties in the sense that the graph size goes to infinity.

We consider graphs with vertex set $V = V_n = \{1, \dots, n\}$ where an edge between the vertices x and y (denoted by $x \sim y$) is interpreted as a persistent contact between the two nodes. Given $x \in V$, its degree is denoted by $d(x)$. We will think of edges as generated by a pair-formation process in which each vertex x - often denoted as an individual - chooses a set of partners according to a specified x -dependent rule. Therefore the set of individuals which have contact with a given vertex x can be divided into two -possible non-disjoint sets: the set of nodes which are chosen by x himself and the set of nodes which have chosen x as one of their partners. We call the size of the first set the outdegree $d_{out}(x)$ of x and the size of the second one the indegree $d_{in}(x)$ of x . Obviously $d(x) \leq d_{out}(x) + d_{in}(x)$ and if the choices are sufficiently independent one can expect equality to hold almost surely if $n \rightarrow \infty$.

We partition the set of vertices V_n into groups $\{C_i(n)\}_{i \geq 1}$ where all members of a group $C_i(n)$ choose exactly i partners by themselves ($d_{out} = i$ on $C_i(n)$). Let $P_\alpha^1(n, j)$ the probability for x to choose a fixed partner $y \in C_j(n)$ if n partners are available for the choice and just one choice will be made be

$$(2.1) \quad P_\alpha^1(n, j) = A_\alpha(n) \frac{j^\alpha}{n}$$

Here $A_\alpha(n)$ is a normalization constant such that $A_\alpha(n) \left(\sum_{i \geq 1} |C_i(n)| \frac{i^\alpha}{n} \right) = 1$ and α is a real parameter. Since we want $A_\alpha(n) \rightarrow A_\alpha$, as $n \rightarrow \infty$ we need $\sum_i |C_i(n)| \frac{i^\alpha}{n}$ to be bounded as a function of n which will impose later on constraints on the constant α . The parameter α acts as an affinity parameter tuning the tendency to choose a partner with a high outdegree or low outdegree. If $\alpha = 0$, choices are made without any preferences and $A_\alpha(n) \equiv 1$. For $\alpha > 0$ the "highly active" individuals are preferred whereas for $\alpha < 0$ the "low activity" individuals are favored. From this we obtain the basic probability

$$\Pr[x \text{ from } C_i(n) \text{ has chosen } y \in C_j(n) \text{ as one of his partners}] \simeq A_\alpha(n) \frac{i \cdot j^\alpha}{n}$$

Concerning the size of the sets $C_i(n)$ we will make the following assumption:

$$(2.2) \quad \frac{|C_i(n)|}{n} =: p_i(n) \xrightarrow{n \rightarrow \infty} \frac{c_1}{i^\gamma}.$$

With this choice we have to impose the restriction $\alpha < \gamma - 1$ to ensure the convergence of $A_\alpha(n)$. We require furthermore $\gamma > 2$ throughout the paper since otherwise the expected indegree for individuals from a fixed group would diverge. The basic probabilities together with the fixed outdegree-distribution define a probability distribution on each graph with vertex-set V_n , and therefore a random graph space $\mathcal{G}_n(\alpha, \gamma)$.

We want to investigate the threshold properties of epidemic processes on these random graphs as a function of the parameters γ and α . Here we are mainly interested in the question whether the epidemics is asymptotically ($n \rightarrow \infty$) sub- or overcritical. For the threshold study in this random graph model it is enough to investigate the epidemic process as a multi-group branching process since typical graphs have locally a tree structure. Absence of a threshold means here that for arbitrary small transition probability $\mu > 0$ one still gets for $n \rightarrow \infty$ an overcritical epidemic process. In section 3 we will compare the results for the above model with some specific graph models which have no power-law-distribution but still diverging second moments of the degree distribution.

First we want to compute the important pairing probabilities which will be used frequently. We start with the easier case $\alpha = 0$.

$$(2.3) \quad \Pr(x \sim y \mid x \in C_i; y \in C_k) = \frac{i+k}{n} - \frac{ik}{n^2} \xrightarrow{n \rightarrow \infty} \frac{i+k}{n}$$

Completely similar one can compute the corresponding probabilities for $\alpha \neq 0$. Dropping the simple details we just state the result:

$$(2.4) \quad \Pr(x \sim y \mid x \in C_i; y \in C_k) \simeq \frac{A_\alpha(ki^\alpha + k^\alpha i)}{n} \text{ as } n \rightarrow \infty.$$

It turns out, that the typical graphs in this model still have for $\alpha < 2$ a *power-law* distribution for the degree with an exponent which can be different from the exponent of the outdegree. For $\alpha > 2$ we obtain a degree distribution which follows in mean a power law but has gaps. To compare both domains we will use the integrated tail distribution $F_k := \Pr(d(x) > k)$. We will show that in both cases we get the same integrated tail distribution. Since we are interested in the dependence of the epidemic threshold from the power-law exponent of the total degree distribution we have to analyze how this exponent varies with the two parameters α and γ . Since the partner choice is sufficiently random and not too strongly biased toward high degree individuals (that's the meaning of the condition $\alpha \leq \gamma - 1$) it is easy to see that the indegree distribution of a vertex from group C_i converges for $n \rightarrow \infty$ to a Poisson distribution with mean $\text{const} \cdot i^\alpha$. There are essentially two regimes in the parameter space, one for which the expected indegree is of smaller order than the outdegree over all groups and one where the indegree is asymptotically of larger order. In the first case it is clear that the indegree is too small to have an effect on the degree distribution exponent. In other words: the set of individuals with degree k consists mainly of individuals whose outdegree is of order k . An easy estimation using the formula for the pairing probabilities shows that the expected indegree of individuals from group i is given by $\mathbb{E}(d_{in}(x) \mid x \in C_i) \simeq \text{const} \cdot i^\alpha$ asymptotically.

Therefore the indegree is of smaller order than the outdegree if $\alpha < 1$. In the case $\gamma - 1 \geq \alpha \geq 1$ the set of individuals with degree k consists mainly of individuals from groups with an index of order $k^{\frac{1}{\alpha}}$. Furthermore we will show that the condition for absence of an epidemic threshold is precisely given by the divergence of the second moment of the degree distribution.

We will state now the main result of the paper which will be shown in the next sections.

2.0.1. *Main Results:* a) for $\alpha < 1$ the random graph $\mathcal{G}_n(\alpha, \gamma)$ has a power-law

distribution for the degree with exponent γ

b) for $1 \leq \alpha < \min(2; \gamma - 1)$ the random graph $\mathcal{G}_n(\alpha, \gamma)$ has a power-law distribution for the degree with exponent $\gamma' = 1 + \frac{\gamma}{\alpha} - \frac{1}{\alpha}$

c) for $2 < \alpha < \gamma - 1$ the random graph $\mathcal{G}_n(\alpha, \gamma)$ has a power-law integrated tail distribution with exponent $\frac{\gamma}{\alpha} - \frac{1}{\alpha}$

d) there is an epidemic threshold for the parameter set $\{\alpha < \frac{\gamma-1}{2}, \gamma > 3\}$ and absence of an asymptotic threshold on the complement

e) for $\alpha > \frac{\gamma-1}{2}$ or $\gamma < 3$ the second moment of the degree distribution diverges.

2.1. The degree distribution and second moment estimation. In this paragraph we want to estimate the degree distribution, respectively the integrated tail distribution for the parameter domains $1 < \alpha < 2$ and $2 < \alpha < \gamma - 1$ and give conditions for the divergence of the second moments. We first note that for $\alpha > 1$ the indegree dominates the outdegree in every group C_i . Furthermore for $n \rightarrow \infty$ and any groupindex i the indegree distribution converges to a Poisson distribution with expectation $\lambda(i) = \text{const} \cdot i^\alpha$. Since the variance is of order i^α it is clear that for $\alpha > 2$ and large i there will be essential no overlap between the indegree distribution in group i and group $i + 1$.

We first deal with the indegree distribution for $\alpha < \min(2; \gamma - 1)$:

$$(2.5) \quad \Pr(d_{in}(x) = k) \simeq \sum_i \Pr(d_{in}(x) = k \mid x \in C_i) \cdot \frac{c_1}{i^\gamma}$$

$$(2.6) \quad = \sum_{i \leq k} \frac{(\text{const} \cdot i^\alpha)^k}{k!} e^{-\text{const} \cdot i^\alpha} \cdot \frac{c_1}{i^\gamma}$$

. The essential interval of indices i which contribute for sufficiently large k to the sum is given by the condition $k \in [\text{const} \cdot i^\alpha - a(k) i^{\frac{\alpha}{2}}; \text{const} \cdot i^\alpha + a(k) i^{\frac{(1+\varepsilon)\alpha}{2}}]$ where $a(k)$ is a slowly growing function in k (for instance $\log k$ is fine). Solving for i and denoting the above constant by \tilde{A} the we get for the boundaries of the essential index interval $I_{ess}(k) := [i_{\min}(k); i_{\max}(k)]$:

$$(2.7) \quad i_{\min}(k) : = \left(-\frac{a(k)}{\tilde{A} \cdot 2} + \left(\frac{a^2(k)}{\tilde{A}^2 \cdot 4} + \frac{k}{\tilde{A}} \right)^{\frac{1}{2}} \right)^{\frac{2}{\alpha}}$$

$$(2.8) \quad i_{\max}(k) : = \left(\frac{a(k)}{\tilde{A} \cdot 2} + \left(\frac{a^2(k)}{\tilde{A}^2 \cdot 4} + \frac{k}{\tilde{A}} \right)^{\frac{1}{2}} \right)^{\frac{2}{\alpha}}$$

. For the length $|I_{ess}(k)|$ of the essential interval we get therefore:

$$(2.9) \quad |I_{ess}(k)| = i_{\max}(k) - i_{\min}(k) = \left(\frac{k}{\tilde{A}}\right)^{\frac{1}{\alpha}} \left(\frac{2a(k)}{\tilde{A}^{\frac{1}{2}}\alpha k^{\frac{1}{2}}} + O\left(\frac{1}{k}\right)\right)$$

. For $\alpha < 2$ it follows that $I_{ess}(k) \cap I_{ess}(k+1) \neq \emptyset$ for all k sufficiently large. In contrast to this we have for $\alpha > 2$ the situation that certain k -values do essentially not appear since $|I_{ess}(k)|$ converges to zero as k tends to infinity and $I_{ess}(k)$ has to contain an integer to give a contribution. We estimate now the sum in the indegree distribution for large k by restricting the summation to the essential intervals:

$$(2.10) \quad \Pr(d_{in}(x) = k) \simeq \sum_{i \in I_{ess}(k)} \frac{(\tilde{A} \cdot i^\alpha)^k}{k!} e^{-\tilde{A} \cdot i^\alpha} \cdot \frac{c_1}{i^\gamma}$$

. Using stirlings formula $k! = (2\pi k)^{\frac{1}{2}} \left(\frac{k}{e}\right)^k (1 + o(1))$ and the approximation $(1 + \frac{\varepsilon}{k})^k = e^{\varepsilon - \frac{\varepsilon^2}{2k} + o(1)}$ for $|\varepsilon| \ll k^{\frac{2}{3}}$ we obtain:

$$(2.11) \quad \Pr(d_{in}(x) = k) \simeq \frac{2a(k)}{\alpha} \cdot k^{\frac{1}{\alpha} - \frac{1}{2}} \cdot \frac{1 + o(1)}{(2\pi k)^{\frac{1}{2}}}$$

$$(2.12) \quad = \frac{1}{k^{1 + \frac{\gamma}{\alpha} - \frac{1}{\alpha} + o(1)}}$$

. We get therefore an asymptotic power law exponent $\gamma' = 1 + \frac{\gamma-1}{\alpha}$ for the indegree distribution which is also the exponent of the total degree distribution in the case $\alpha > 1$ since then $\gamma' < \gamma$ and hence the indegree dominates the outdegree.

It is much more easy to estimate the integrated tail distribution. We will do this again for the indegree distribution and $\alpha > 2$ to be able to compare with the situation for smaller values of α . Clearly one has

$$(2.13) \quad F_k = \Pr(d_{in}(x) > k) = \sum_i \sum_{x \in C_i} \sum_{l>k} \Pr(d_{in}(x) = l \mid x \in C_i)$$

$$(2.14) \quad \simeq \sum_i \sum_{l>k} \Pr(d_{in}(x) = l \mid x \in C_i) \frac{c_1}{i^\gamma}$$

. Since for the degree $l > k$ only vertices with indices $i > \left\lfloor k^{\frac{1}{\alpha}} (1 + o(k^{\frac{1}{\alpha}})) \right\rfloor$ contribute and in this case the sum $\sum_{l>k} \Pr(d_{in}(x) = l \mid x \in C_i)$ is essentially 1, we get

$$(2.15) \quad F_k = \sum_{i>\left\lfloor k^{\frac{1}{\alpha}} (1 + o(k^{\frac{1}{\alpha}})) \right\rfloor} \sum_{l>k} \Pr(d_{in}(x) = l \mid x \in C_i) \frac{c_1}{i^\gamma}$$

$$(2.16) \quad = \sum_{i>\left\lfloor k^{\frac{1}{\alpha}} (1 + o(k^{\frac{1}{\alpha}})) \right\rfloor} \frac{c_1}{i^\gamma}$$

$$(2.17) \quad = \frac{c_1}{\left\lfloor k^{\frac{1}{\alpha}} (1 + o()) \right\rfloor^{\gamma-1}} = \frac{c_1}{k^{\frac{\gamma}{\alpha} - \frac{1}{\alpha}} (1 + o(k^{\frac{\gamma}{\alpha} - \frac{1}{\alpha}}))}$$

. Therefore the integrated tail-degree distribution is still of power-law form.

Next we want to estimate the second moments M_2^{in} of the indegree distribution for values $1 < \alpha < \gamma - 1$. By definition we have

$$(2.18) \quad M_2^{in} = \sum_k \sum_i \sum_{x \in C_i} k^2 \Pr(d(x) = k \mid x \in C_i)$$

$$(2.19) \quad = \sum_k \sum_i k^2 \Pr(d(x) = k \mid x \in C_i) \frac{c_1}{i^\gamma}$$

and asymptotically since the second moment of a Poisson distribution with expectation λ is given by $\lambda^2 - \lambda$:

$$(2.20) \quad M_2^{in} \simeq \sum_i \frac{c_1}{i^\gamma} \cdot \left[(\tilde{A}i^\alpha)^2 - \tilde{A}i^\alpha \right]$$

$$(2.21) \quad = \sum_i \left(\frac{\text{const}}{i^{\gamma-2\alpha}} - \frac{\text{const}}{i^{\gamma-\alpha}} \right)$$

. The first term in the sum diverges for $\alpha > \frac{\gamma-1}{2}$ which is of course also the condition for the total degree second moment to diverge (in the case $\alpha < 1$ the outdegree dominates the indegree and the second moment diverges for $\gamma < 3$). We show in the next paragraph that this is precisely the condition for absence of an epidemic threshold.

2.2. The branching process approximation. In the following we want to analyze the branching process approximation for the epidemic process on the above described random graph space $\mathcal{G}_n(\alpha, \gamma)$. Due to the absence of local cycles (in the large n limit) one gets an exact threshold estimation via the branching process approximation. Let μ be the transmission probability that an infection will be transmitted along a given edge of the graphs. Eliminating edges in $\mathcal{G}_n(\alpha, \gamma)$ with probability $1 - \mu$ gives a new random graph space $\mathcal{G}_n^\mu(\alpha, \gamma)$ which we call the epidemic random graph. We will denote expectations with respect to $\mathcal{G}_n^\mu(\alpha, \gamma)$ by \mathbb{E}_μ whereas \mathbb{E} is reserved for expectations with respect to $\mathcal{G}_n(\alpha, \gamma)$.

Let $T(\alpha, \gamma, n) = (a_{ij}(\alpha, n))$ be given by

$$(2.22) \quad a_{ij}(\alpha, n) := \mathbb{E}_\mu(\#\{y : x \sim y \mid x \in C_i(n), y \in C_j(n)\})$$

$$(2.23) \quad = \mu \cdot \mathbb{E}(\#\{y : x \sim y \mid x \in C_i(n), y \in C_j(n)\})$$

and let $a_{ij}(\alpha) := \lim_{n \rightarrow \infty} a_{ij}(\alpha, n)$. Note that $T(\alpha, \gamma, n)$ is the transposed of the branching process transition matrix. In other words $a_{ij}(\alpha)$ is the expected number of infected individuals in group j generated by one infected individual in group i . Actually the exact terms in the asymptotic branching process matrix are given by $a_{ij}(\alpha) - 1$. Since the difference by one is irrelevant for our considerations we neglect this term in the computations. From formula (2.4) we conclude that

$$(2.24) \quad a_{ij}(\alpha, n) \simeq C_1 \cdot \mu \cdot \frac{ij^\alpha + ji^\alpha}{j^\gamma}$$

where the constant C_1 depends only on α and γ . Absence of an epidemic threshold for fixed α and γ holds exactly if

$$(2.25) \quad \lim_{n \rightarrow \infty} \lambda_{\max}(T(\alpha, \gamma, n)) = \infty \quad \text{for } \mu > 0$$

with $\lambda_{\max}(\cdot)$ being the largest eigenvalue in modulus.

The regions in the parameter space for which we will get absence of an asymptotic epidemic threshold are 1) $2 < \gamma \leq 3$ and arbitrary α and 2) the parameters with $\gamma > 3$ and $\alpha \geq \frac{\gamma-1}{2}$. Finally we get for $\gamma > 3$ and $\alpha < \frac{\gamma-1}{2}$ that there is a size independent threshold.

Since the smallest row sum of a positive matrix gives a lower bound on the largest eigenvalue it is enough to show that for $(b_{lk}(n)) := T^2(\alpha, \gamma, n)$ and α and γ as in case 1) or 2) we have

$$(2.26) \quad \min_l \sum_{k \geq 1} b_{lk}(n) \xrightarrow{n \rightarrow \infty} \infty.$$

Note that $a_{ij}(\alpha, n) \simeq a_{ij}(\alpha)$ and therefore

$$(2.27) \quad b_{lk}(n) \simeq \sum_i a_{li}(\alpha) a_{ik}(\alpha)$$

$$(2.28) \quad = (\mu C_1)^2 \sum_i \frac{li^\alpha + il^\alpha}{i^\gamma} \cdot \frac{ik^\alpha + ki^\alpha}{k^\gamma}$$

$$(2.29) \quad = (\mu C_1)^2 \left(\frac{lk^\alpha + l^\alpha k}{k^\gamma} \sum_i \frac{i^{\alpha+1}}{i^\gamma} + \right.$$

$$(2.30) \quad \left. + \frac{lk}{k^\gamma} \sum_i \frac{i^{2\alpha}}{i^\gamma} + \frac{l^\alpha k^\alpha}{k^\gamma} \sum_i \frac{i^2}{i^\gamma} \right).$$

. From this formula one immediately concludes the absence of an epidemic threshold in the above described domains since $\sum_i \frac{i^2}{i^\gamma}$ respectively $\sum_i \frac{i^{2\alpha}}{i^\gamma}$ diverge for $\gamma \leq 3$ respectively $\alpha \geq \frac{\gamma-1}{2}$ whereas the divergence of $\sum_i \frac{i^{\alpha+1}}{i^\gamma}$ is irrelevant since $\frac{\gamma-1}{2} \geq \gamma - 2$ for $\gamma \geq 3$.

To show the existence of an asymptotic threshold for parameters satisfying $\gamma > 3$ and $\alpha < \frac{\gamma-1}{2}$ we will demonstrate that for $\mu > 0$ we have $\lim_{m \rightarrow \infty} \lambda_{\max}(T^m(\alpha, \gamma, n)) < \infty$ from which the assertion then follows. Let $(g_{lk}^m(n)) := T^m(\alpha, \gamma, n)$. We will derive a recursion for $g_{lk}^m(n)$ as follows. Since

$$(2.31) \quad g_{lk}^m(n) \simeq (\mu C_1)^m \sum_{i_m \dots i_1} \frac{li_m^\alpha + i_m l^\alpha}{i_m^\gamma} \cdot \dots \cdot \frac{i_2 i_1^\alpha + i_1 i_2^\alpha}{i_1^\gamma} \cdot \frac{i_1 k^\alpha + ki_1^\alpha}{k^\gamma}$$

we define variables

$$(2.32) \quad g_{lk}^m(n, x, y) := (\mu C_1)^m \sum_{i_m \dots i_1} \frac{li_m^\alpha + i_m l^\alpha}{i_m^\gamma} \cdot \dots \cdot \frac{i_2 i_1^\alpha + i_1 i_2^\alpha}{i_1^\gamma} \cdot \frac{i_1 x + yi_1^\alpha}{k^\gamma}$$

and observe that

$$(2.33) \quad g_{lk}^{m+1}(n, x, y) = \mu C_1 \cdot g_{lk}^m(n, x', y')$$

where

$$(2.34) \quad x' = B_0 x + B_1 y ; \quad y' = B_2 x + B_o y$$

and $B_0 = \sum_i \frac{i^{\alpha+1}}{i^\gamma}$, $B_1 = \sum_i \frac{i^{2\alpha}}{i^\gamma}$ and $B_2 = \sum_i \frac{i^2}{i^\gamma}$. Since $g_{lk}^m(n, k^\alpha, k) \simeq g_{lk}^m(n)$ we only have to show that $\mu C_1 \cdot \lambda_{\max} \left[\begin{pmatrix} B_0 & B_1 \\ B_2 & B_o \end{pmatrix} \right] < 1$ to establish the boundedness of $\lambda_{\max}(T^m(\alpha, \gamma, n))$. Clearly if B_0, B_1 and B_2 stay bounded as a function of n we

can match the requirement of the last sentence. But B_0, B_1 and B_2 stay bounded on the parameter set $\gamma > 3$ and $\alpha < \frac{\gamma-1}{2}$ from which our assertion follows.

The result has a nice heuristic explanation. The "tail"-groups (that is the groups C_i with i very large) act as a hub for the infection. Indeed given an infected individual in group C_k it spreads the infection into the hub via its outdegree edges (for α large) from which the infection is backspread via the indegree edges of the hub to C_k or any other fixed group. It is precisely this infection path which is proportional to $\sum_i i^{2\alpha}$ and therefore makes the branching process divergent for $\alpha > 2\gamma + 1$.

3. THE VERY SMALL WORLD INSIDE THE SMALL WORLD

In the following we will show that the distinction between absence and presence of an epidemic threshold can be exactly related to properties of the diameter of the giant component $\mathcal{G}_n^0(\alpha, \gamma)$ of $\mathcal{G}_n(\alpha, \gamma)$. Of course the diameter of a random graph space is itself a random variable but it turns out that it has a small variation. For our purpose we therefore concentrate on the expected diameter. Given a graph G_n with n vertices and a subset A of V_n we define $diam(G_n, A) := \max \{d(x, y) \mid x, y \in A\}$ where $d(x, y)$ is the distance on the graph G_n . Let the ε -essential diameter of a graph G_n be defined as $diam_\varepsilon^{ess}(G_n) := \min_A \{diam(G_n, A) \mid |A| \geq (1 - \varepsilon)|V_n|\}$. Similar if we have a sequence of random graph spaces $\{\mathcal{G}_n\}$ the quantity $diam_\varepsilon^{ess}(\mathcal{G}_n)$ is now a random variable depending on the realization $G_n \in \mathcal{G}_n$.

The striking phenomenon in the case of absence of an epidemic threshold for the above described parameter region is the following property:

For all $\varepsilon \in (0, 1)$ there is a $C(\varepsilon) > 0$ such that $\Pr \{diam_\varepsilon^{ess}(\mathcal{G}_n^0(\alpha, \gamma)) < C(\varepsilon)\} \rightarrow 1$ as n tends to infinity.

In other words: A positive fraction of the vertices are located in a cluster of finite diameter. This cluster has a kind of fuzzy, onion like structure with no sharp boundary. At the "center" of the cluster are the very high degree groups placed which are entirely contained in the cluster. The union of shells which are at distance less than R from the center has positive mass as soon as R gets larger than a critical value C_0 . Increasing R makes this fraction larger and larger but to get the whole giant connected component in the cluster requires to take $R \rightarrow \infty$ as n goes to infinity (something between $\log \log n$ and $\log n$ should be the right growth rate).

Before going into the details let us recall the general philosophy how to estimate the diameter of a (large) connected component of a random graph space. Fixing an individual x one tries to estimate the expectation of $\Gamma_k := \#\{y : d(x, y) = k\}$ as good as possible. As long as $\sum_{i \leq k} \Gamma_i \ll n$ holds and the variation of the Γ_k is under control one can use in general a branching process approximation to get the right order of the expectation of Γ_k . The smallest k_0 such that $\sum_{i \leq k_0} \Gamma_i \sim n$ gives then a very good estimation of $diam_\varepsilon^{ess}$ (for ε small).

To see that the above statement is true let us note first that iterating the branching process matrix k -times -denoted in the following by $(t_{ij}^{(k)})$ and being the transposed of $T^k(\alpha, \gamma)$ —gives the expected size of the number of individuals in group i which are at distance k to a random chosen individual in group j provided that all expectations $\sum_j t_{ij}^{(k')}$ are much smaller than $|C_i(n)|$ for $k' \leq k$.

Assume now that the maximal outdegree respectively the largest group index in our model scales like n^β for some $0 < \beta < \frac{1}{\gamma}$ and hence the size of the group with the maximal outdegree is of order $\text{const} \cdot n^{1-\beta\gamma}$. Since the dominating contribution in the $t_{ij}^{(k)}$ terms comes from $\left(\sum_{l < n^\beta} \frac{\text{const}}{l^{\gamma-2\alpha}}\right)^k$ we can safely iterate the branching process matrix k -times as long as $k\beta(1-\gamma+2\alpha) < 1 - \beta\gamma$ (note that in the "no-threshold" case we always have $1 - \gamma + 2\alpha > 0$). Let k_0 be the largest k for which the above inequality holds. To get a good lower bound on the number of individuals at distances larger than k_0 one has to truncate the matrix properly in such a way that only paths are taken into account which make no use of groups which are smaller than the entries in the iterated branching process matrix. Therefore we consider in step $k_0 + k$ only those groups C_i for which $i < n^{\beta_k}$ with $\beta_k := \frac{1}{(1-\gamma+2\alpha)(k_0+k)+\gamma}$. The leading term after $k_0 + k$ iterations is now given by $n^{k_0\beta(1-\gamma+2\alpha)} \prod_{l \leq k} n^{\beta_l(1-\gamma+2\alpha)} = n^{k_0\beta(1-\gamma+2\alpha) + \sum_{l \leq k} \frac{(1-\gamma+2\alpha)}{(1-\gamma+2\alpha)(k_0+k)+\gamma}}$. Since the sum in the exponent diverges there is a size independent k_1 such that $n^{k_0\beta(1-\gamma+2\alpha)} \prod_{l \leq k_1} n^{\beta_l(1-\gamma+2\alpha)} \simeq n$ and consequently a finite fraction of the vertex set stays within a bounded distance.

Let us finally demonstrate that there cannot be a C^* such that essentially all vertices of the giant component are within a distance less than C^* . Fix $f > 0$ and an individual x with $d(x) < f$. From formula (2.4.) it follows that there is a n independent $p(f) > 0$ such that the probability that x has only partners y with $d(y) < f$ is larger $p(f) > 0$. Therefore the probability that at distance less equal C^* from x there are only individuals with degree less than f is larger than $p(f)^{f^{C^*+1}}$. Since there are $\text{const}(f) \cdot n$ individuals with degree less than f there is a small but positive fraction of individuals which are at distance larger than C^* from the high degree groups for arbitrary large C^* . In a similar way one can show that there are individuals with distance of order $\ln \ln n$ (but those form only a set of asymptotic zero density).

4. A TOY EXAMPLE

In the following section we want to discuss a model in which the degree-distribution is somehow extreme, namely there are only two types of individuals with sizes C_2 and $C_{f(n)}$. The index 2 and $f(n)$ stands for the degree of the two individual-types the last one depends on the size n of the vertex set. We will make the general requirement that the second moment of the degree distribution diverges whereas the first moment converges:

$$(4.1) \quad \limsup f(n) \frac{|C_{f(n)}|}{n} < \infty$$

$$(4.2) \quad [f(n)]^2 \frac{|C_{f(n)}|}{n} \rightarrow \infty.$$

We will look at the epidemic process initiated by one randomly infected individual in group $C_{f(n)}$ and ask whether one has a persistent epidemics for arbitrary small transmission probability. For this model the question of existence of an epidemic

threshold is only interesting in the above conditional sense since due to our assumptions the asymptotic fraction of $C_{f(n)}$ individuals on the total population tends to zero and no epidemics can be persistent among the C_2 population for sufficiently small transmission probabilities.

We will investigate two different structures of edge formations.

4.1. Case A. In this subsection we want to separate the elements of $C_{f(n)}$ as far as possible from each other which gives the epidemiologically "safest" structure in the sense that unless the first moments of the degree-distribution "almost" diverge one would expect an undercritical situation if the transmission probability is sufficiently small (note that if the first moment of the degree-distribution diverges one gets trivially no threshold). We want to separate all elements of $C_{f(n)}$ by chains of uniform length $l(n)$ consisting of elements of C_2 . Since we have $f(n) \frac{|C_{f(n)}|}{2}$ chains (we drop all effects due to integer part requirements since they are asymptotically irrelevant) we get for the length $l(n) \cong \frac{2|C_2|}{|C_{f(n)}|f(n)} + 1$. The probability to transmit an infection along such a chain is $\mu^{l(n)}$ and hence the expected number of individuals in $C_{f(n)}$ which will be infected from one initially infected individual from group $C_{f(n)}$ equals $f(n) \mu^{l(n)}$. To get absence of a threshold we need therefore $f(n) \mu^{l(n)} > 1$ asymptotically. This is equivalent to $\log f(n) - c \frac{(n - |C_{f(n)}|)}{|C_{f(n)}|f(n)} > 0$ with $c = 2 \log \frac{1}{\mu}$ from which by (4.1) $O(n) \log f(n) > c \cdot (n - |C_{f(n)}|)$ follows. First we observe that it is impossible to find exponents $\beta, \delta \in (0, 1)$ s.t. $f(n) := n^\beta$ and $|C_{f(n)}| := n^\delta$ and $f(n) \mu^{l(n)} \rightarrow \infty$ under the assumption $\beta + \alpha < 1$. Note that $\beta + \alpha = 1$ is just the borderline where the expectation of the degree starts to become large.

As a conclusion from the considerations in this section one should probably take, that it is the balance between separation length and degree growth which matters for the epidemiological threshold-question. So if the elements from C_2 would not be as uniformly spread we should look at the following quantities: $\varphi_l[f(n)] := \mathbb{E}_x [\#(y \in C_{f(n)} : dist_0(x, y) = l)]$ where $dist_0$ denotes the length of the shortest path between $x \in C_{f(n)}$ and y which entirely consists of individuals from C_2 . For getting an overcritical epidemics we therefore need $A(n) := \sum_{l \geq 0} \varphi_l[f(n)] \mu^l > 1$.

4.2. Case B. Assume now that we have stationary asymptotic probabilities p_{00} , p_{01} , p_{10} and p_{11} for the events that a single random chosen edge from an $x \in C_{f(n)}$ points to an element from $C_{f(n)}$ (p_{00}) or to an element from C_2 (p_{01}) respectively that an edge from an element from C_2 points to $C_{f(n)}$ (p_{10}) or to an element from C_2 (p_{11}). There are three conditions on the choice of the p_{ij} namely $p_{00} + p_{01} = 1$, $p_{10} + p_{11} = 1$ and $p_{01} \cdot f(n) |C_{f(n)}| = p_{10} \cdot 2|C_2|$. Since the first moments are bounded we have $|C_{f(n)}| = o(n)$ and therefore $|C_2| \sim n$. To get the threshold condition we have to estimate the expected number of secondary infections caused by an infected individual from $C_{f(n)}$ which is precisely given by the quantity $A(n)$. To compute the numbers $\varphi_l[f(n)]$ observe that

$$(4.3) \quad \Pr(\text{a fixed chain starting at } x \text{ has length } l, l > 1) = p_{01} p_{11}^{l-2} p_{10}$$

and $\Pr(\text{a fixed chain starting at } x \text{ has length 1}) = p_{00}$. Therefore we have

$$(4.4) \quad \sum_{l \geq 1} \varphi_l [f(n)] \cdot \mu^l = f(n) \left(\mu p_{00} + \sum_{l > 1} \mu^l p_{01} p_{11}^{l-2} p_{10} \right)$$

$$(4.5) \quad = f(n) \mu p_{00} + f(n) \mu^2 p_{01} p_{10} \frac{1}{1 - \mu p_{11}}$$

(we dropped the n -dependence of the p_{ij}). Clearly the direct coupling between the $C_{f(n)}$ individuals has to be of order at most $\frac{1}{f(n)}$ to avoid a nontrivial epidemics inside the $C_{f(n)}$ population which implies $f(n)|C_{f(n)}| \ll C_2$ and in turn $p_{10} \rightarrow 0$. Using the relations between the p_{ij} we have $A > \mu^2 \cdot \frac{f^2(n)|C_{f(n)}|}{2|C_2|}$. From the divergence of the second moments one gets $\frac{f^2(n)|C_{f(n)}|}{2|C_2|} \rightarrow \infty$ and hence absence of an epidemic threshold in this probabilistic setting. It remains to estimate the asymptotic number $I(n)$ of infected individuals which equals the size of the largest component of the epidemic random graph. We first estimate the number of infected individuals in $C_{f(n)}$ denoted by I_f via the stationarity equation $|C_{f(n)}| \left(1 - \frac{1}{A(n)}\right) = I_f(n)$. Furthermore an infected individual from $C_{f(n)}$ infects in mean $\sum_{l \geq 1} f(n) \cdot \mu^l = f(n) \frac{1}{1-\mu}$ individuals from C_2 from which one obtains the following asymptotic bounds on the number of infected individuals: $\frac{1}{2} |C_{f(n)}| \left(1 - \frac{1}{A(n)}\right) f(n) \frac{1}{1-\mu} \leq I(n) \leq |C_{f(n)}| \left(1 - \frac{1}{A(n)}\right) f(n) \frac{1}{1-\mu}$. Therefore unless one has $\lim_{n \rightarrow \infty} \frac{|C_{f(n)}| f(n)}{|C_2|} > 0$ (in this case one gets trivially absence of a threshold since $p_{00} > 0$) the size of the asymptotic fraction of infected individuals in the case of absence of a threshold still converges to zero.

5. CONCLUSIONS AND COMMENTS

We have shown that for the preferential choice model described in this paper the divergence of the second moment is precisely the condition for absence of an epidemic threshold even in the case when the asymptotic distribution is not of a power-law form (but has gaps). In terms of graph geometry this is related to the presence or absence of a massive finite diameter cluster which carries the epidemics. It seems that for the case of an exponent larger than three one always has a size independent threshold if one additionally requires that the size of the largest component of the epidemic random graph is proportional to $\mu \cdot n$. Note that it is easy to get absence of threshold for exponents larger than three if one introduces horizontal preferences in the partner choice (for instance a strong bias to make connections into the own group) but in this case the size of the largest component scales like $\mu^{1+\epsilon} \cdot n$ with an $\epsilon > 0$. It is a much more difficult task to estimate exactly the size of the giant component on the epidemic random graph which will be the subject of a more mathematical forthcoming paper [6]. Certainly it scales like $\text{const} \cdot \mu \cdot n$. It would be furthermore very interesting to have random network growth principles which produce in a natural way the scale free graphs we described in this paper. For some results in this direction see [10]. In any case the model seems to be interesting in describing computer networks and associated spread of viruses on it. Caution is required if one wants to draw conclusions for sexually transmitted diseases like HIV on social networks even when these networks have

a power law tail distribution for the degree. The main point here is that the transmission probability scales with the degree for small transmission probabilities-like in the AIDS-epidemics- since high degree individuals spend necessarily less time with each single partner [8].

Finally we would like to mention that there is a natural variant of our model by defining independent directed edge probabilities between $x \in C_i$ and $y \in C_j$ as $P_{ij} = A_\alpha \frac{l^{\alpha} i}{n}$ with the same A_α as above. For this model one gets the same threshold statements since it has the same branching process transition matrix. For some alternative strategies to choose partners and the corresponding threshold estimations see also[11][5].

REFERENCES

- [1] R.Albert, A.-L.Barabasi: *Statistical Mechanics of Complex Networks*, Reviews of Modern Physics, **74**, 47 (2002), arXiv:cond-mat/0106096
- [2] R.Albert, A.-L.Barabasi: Science **286**, 509 (1999)
- [3] P.Bak, C.Tang, K.Wiesenfeld: *Self organized criticality: An explanation of 1/f noise*, Phys.Rev.Lett., **59**, 381 (1987)
- [4] P.Bak, *How Nature Works*, Springer-Verlag (1996)
- [5] Ph.Blanchard, G.-F.Bolz, T.Krüger: *Modelling AIDS-Epidemics or any venereal diseases on random graphs*, in "Stochastic Processes in Epidemic Theory", Lecture Notes in Biomathematics **86**, Springer-Verlag (1990)
- [6] Ph.Blanchard, C.-H.Chang, T.Krüger, St.Reiman: *Epidemic thresholds on scale-free graphs*, BiBoS Preprint, University of Bielefeld (2002)
- [7] B.Bollobas, O.Riordan, J.Spencer, G.Tusnady: *The degree sequence of a scale-free random graph*, Random Struc.Alg., **18**, 279 (2001)
- [8] Z.Deszö, A.-L.Barabasi: *Halting viruses in scale-free networks*, arXiv:cond-mat/0107430v2 (2002)
- [9] R.Pastor-Satorras, A.Vespignani: *Epidemic Spreading in Scale-Free Networks*, Phys.Rev.Lett.**86**, 3200 (2001)
- [10] D.Volchenkov, Ph.Blanchard: *An algorithm generating random graphs with power law distributions*, to appear in Physica A (2003)
- [11] D.Volchenkov, L.Volchenkova, Ph.Blanchard: *The role of social structure and immunization policy in disease spreading in scale free networks*, submitted to Phys.Rev.E

A.ONE: UNIVERSITÄT BIELEFELD, FAKULTÄT FÜR PHYSIK, 33619 BIELEFELD, UNIVERSTÄTSSTR.25, GERMANY—, A.TWO: NATIONAL CENTER FOR THEORETICAL SCIENCES, PHYSICS DIVISION, NTHU 101.SECTION2, HSINCHU, TAIWAN 300, R.O.C.—, A.THREE: UNIVERSITÄT BIELEFELD, FAKULTÄT FÜR MATHEMATIK, 33619 BIELEFELD, UNIVERSITÄTSSTR.25, GERMANY, TKRUEGER@PHYSIK.UNI-BIELEFELD.DE—

E-mail address: blanchard@physik.uni-bielefeld.de, chchang@phys.cts.nthu.edu.tw, tkrueger@physik.uni-bielefeld.de